Nonparametric Transformation to White Noise

Oliver B. Linton* — London School of Economics

E. Mammen† — Universität Mannheim

PRELIMINARY AND INCOMPLETE

September 30, 2005

Abstract

We propose a new estimator of a nonparametric regression subject to time series errors that improves on Xiao et al. (2003). Our method is based on a different whitening transformation that produces a type 2 linear integral estimating equation for the regression function. We investigate both the stationary case and the case where the error has a unit root. In the stationary case we achieve efficiency improvements. In the unit root case our procedure is consistent and asymptotically normal unlike the standard regression smoother. We also present the distribution theory for the parameter estimates, which is non-standard in the unit root case. We also investigate its finite sample performance and demonstrate its effectiveness.

Key words: Efficiency; Inverse Problem; Kernel Estimation; Nonparametric regression; Unit Roots.

Journal of Economic Literature Classification: C14

1 Introduction

In this paper we discuss the estimation of the unknown quantities in the model

$$B(L)Y_t = A(L)m(X_t) + \varepsilon_t, \tag{1}$$

*Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: lintono@lse.ac.uk. Supported by the Cowles Foundation and the Economic and Social Science Research Council.

†Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany. E-mail address: emammen@rumms.uni-mannheim.de. Tel. 0049 621 181 1927. Supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse", Humboldt-Universität zu Berlin and Project MA1026/6-2.

where ε_t is a martingale difference sequence and mean independent of the regressors X_t, while $A(L) = \sum_{j=0}^{\infty} a_j L^j$ and $B(L) = \sum_{j=0}^{\infty} b_j L^j$ are lag polynomial operators with $a_0 = b_0 = 1$, where $Lx_t = x_{t-1}$. The function $m(.)$ is assumed to be unknown but smooth, and is the object of central interest, although the dynamics of the model represented by $A(L), B(L)$ are also fundamental to the interpretation. We treat only the case where $A(L), B(L)$ are described by a finite dimensional parameter $\theta = (\alpha, \beta) \in \mathbb{R}^p$ with $\alpha \in \mathbb{R}^{p_a}$ parameterizing A and $\beta \in \mathbb{R}^{p_b}$ parameterizing B. There are two main cases to consider: (a) both Y_t and X_t stationary and short memory; (b) either X_t or Y_t or both are non-stationary or long memory.

In the stationary case (a) the main issue is efficiency. A special case of interest is the nonparametric regression model

$$Y_t = m(X_t) + u_t,\ t = 1, \ldots, T, \tag{2}$$

where the covariates follow some stationary mixing process, while the residual process u_t satisfies

$$A(L)u_t = \varepsilon_t = \sum_{j=0}^{\infty} a_j u_{t-j}. \tag{3}$$

In this case, $A(L)Y_t = A(L)m(X_t) + \varepsilon_t$, which is a special case of (1) with $A(L) = B(L)$. In this model there are many standard estimators of m and of the parameters of $A(L)$ that are consistent. However, unlike in the parametric case, the standard kernel regression smoothers do not take account of the correlation structure in X_t or u_t and estimate the regression function in the same way as if these processes were independent. Furthermore, the variance of such estimators is proportional to the short run variance of u_t, $\sigma_u^2 = \mathrm{var}(u_t)$ and does not depend on the regressor or error covariance functions $\mathrm{cov}(X_t, X_{t-j})$, $\mathrm{cov}(u_t, u_{t-j})$, $j \neq 0$. Practitioners accustomed to correcting standard errors for dependence believe that the standard errors in nonparametric regression are therefore suspect. As Conley, Hansen, Luttmer, and Scheinkman (1997) say: "Although theoretically correct the practice of ignoring serial correlation is not likely to work well for the temporal dependence present in our short-term interest rate data". This point has been addressed recently by Xiao, Linton, Carroll, and Mammen (2003) who proposed a more efficient estimator of m based on a prewhitening transformation

$$Y_t - \sum_{j=1}^{\infty} a_j (Y_{t-j} - m(X_{t-j})) = m(X_t) + \varepsilon_t,$$

where the right hand side is now a standard nonparametric regression with whitened errors (and replacing the unknown quantities on the left hand side by preliminary estimates of m and α). The transform implicitly takes account of the autocorrelation structure. They obtained an improvement in terms of variance over the usual kernel smoothers.

We propose an alternative strategy for estimation of m along with the parameters of $A(L)$ in (2, 3). This is essentially to estimate the transformed model (1) as an additive (possibly infinite order) nonparametric regression. Recently, Linton and Mammen (2005) have shown how to estimate similar models using the theory of linear integral equations of the second kind; see also Carrasco, Florens and Renault (2002). We obtain an estimating equation for m that is a type 2 linear integral equation for each parameter value θ. To obtain the parameters θ we optimize a profile likelihood criterion. We show that our method has attractive theoretical and finite sample properties. In particular, it has smaller asymptotic variance than the main method of Xiao, Linton, Carroll, and Mammen (2003) and furthermore the asymptotics require weaker conditions. The parametric version of the regression model (2) and (3) is a standard teaching topic in graduate econometrics, Harvey (1981, Chapter 6). The traditional applications were in for example production studies where Y_t is output and X_t is the capital/labour ratio of a given firm or industry observed over time. What is of interest is the function m and its derivatives and it is not essential that the error term be serially uncorrelated. In fact in many parametric studies serial correlation has been found in error terms.

We define our method in the more general model (1). The more general model (1) allows for richer dynamics and is more plausible, see Harvey (1981, Chapter 7). For example, it is consistent with a very general linear partial adjustment mechanism of actual Y to desired Y^* when $Y^* = m(X)$. It also corresponds more directly to the general ARMAX class of models treated in Hannan and Deistler (1988) except that we have a particular nonlinear component.

We also consider the case (b) where some of the variables are nonstationary. This could arise for example from a unit root in the residual u_t or in X_t or in both, see Phillips and Park (1998). In this case, estimating in the original data (2) may lead to inconsistency, whereas the transformation involved in (1) yields error terms with a lower order of nonstationarity/persistence and hence consistency can be obtained. The estimation method is more or less the same as in the stationary case although the justification of it differs. The distribution theory for the parametric part though is non standard in this case: in fact we obtain T convergence to Dickey-Fuller distributions under the unit root.

2 The Stationary Case

In this section we suppose that (Y_t, X_t) are jointly stationary and weakly dependent mixing processes.

2.1 Estimation Method

2.1.1 Population Characterization

We first suppose that $A(L), B(L)$ are known. Letting $Z_t = B(L)Y_t$ we have

$$Z_t = A(L)m(X_t) + \varepsilon_t = \sum_{j=0}^{\infty} a_j m(X_{t-j}) + \varepsilon_t,$$

which is an additive autoregression with i.i.d. errors where the additive components are subject to the restriction that they all share a common function m. In view of the assumed stationarity, define the function m as the minimizer of the criterion

$$Q(\theta_0, m) = E\left[\left\{Z_t - \sum_{j=0}^{\infty} a_j m(X_{t-j})\right\}^2\right]. \tag{4}$$

A necessary condition for m to be the minimizer is that it satisfies the first order condition

$$E\left[\left\{Z_t - \sum_{j=0}^{\infty} a_j m(X_{t-j})\right\} \sum_{k=0}^{\infty} a_k h(X_{t-k})\right] = 0 \tag{5}$$

for any measurable function h. This implies, taking $h(.)$ to be the Dirac delta function, that

$$\sum_{j=0}^{\infty} a_j E[Z_t | X_{t-j} = x] = \sum_{j=0}^{\infty} a_j^2 m(x) + \sum\sum_{j \neq k} a_j a_k E[m(X_{t-j}) | X_{t-k} = x]. \tag{6}$$

This is an implicit equation for $m(.)$. It can be re-expressed as a linear type 2 integral equation in $L_2(f_0)$, where f_0 is the marginal density of X_t. Define $a_j^* = a_j / \sum_{j=0}^{\infty} a_j^2$ and $a_j^+ = \sum_{k \neq 0} a_{j+k} a_j / \sum_{l=0}^{\infty} a_l^2$, and let $f_{0,j}$ be the joint density of (X_t, X_{t-j}) and f_0 be the marginal density of X_t. Then

$$m(x) = m^*(x) + \int \mathcal{H}(x, y) m(y) f_0(y) dy, \text{ or } m = m^* + \mathcal{H}m, \tag{7}$$

$$m^*(x) = \sum_{j=0}^{\infty} a_j^* E[Z_t | X_{t-j} = x]$$

$$\mathcal{H}(x, y) = -\sum_{j=\pm 1}^{\pm\infty} a_j^+ \frac{f_{0,j}(y, x)}{f_0(y) f_0(x)}.$$

This is similar to the equation derived in Linton and Mammen (2005) with the exception that there X_t was lagged values of Y_t. Equation (7) is an implicit equation in m and we need some conditions on the operator $\mathcal{H}(x, y)$ to guarantee that there exists a unique solution.

ASSUMPTION A1. *The operator $\mathcal{H}(x, y)$ is Hilbert-Schmidt i.e.,*

$$\int\int \mathcal{H}(x, y)^2 f_0(x) f_0(y) dx dy < \infty.$$

A sufficient condition for A1 is that the joint densities $f_{0,j}(y, x)$ have compact support and are bounded away from zero on this support, which we shall assume below

Under assumption A1, $\mathcal{H}$ is a self-adjoint bounded compact linear operator on the Hilbert space of functions $L_2(f_0)$, and therefore has a countable number of eigenvalues[1]:

$$\infty > |\lambda_1| \geq |\lambda_2| \geq \ldots,$$

[1] These are real numbers for which there exists functions $e_j(.)$ such that $\mathcal{H}e_j = \lambda_j e_j$.

with $\sum_{j=0}^{\infty} \lambda_j^2 < \infty$.

ASSUMPTION A2. *There exist no* $m \in \mathcal{M}$ *with* $\|m\|_2 = 1$ *such that* $\sum_{j=0}^{\infty} a_j m(X_{t-j}) = 0$ *with probability one.*

This condition rules out a certain 'concurvity' in the stochastic process. That is, the data cannot be functionally related in this particular way. In the AR(1) case this says that there are no functions m with $\|m\|_2 = 1$ that satisfy $m(X_t) - \rho m(X_{t-1}) = 0$ with probability one.

Under A1-A2 there exists a unique solution to (7) that satisfies

$$m = (I - \mathcal{H})^{-1} m^*. \tag{8}$$

This is the main characterization used for estimation, although we must first extend this to the case where a general θ is used not necessarily the true θ_0.

For each β, θ, define $Z_t(\beta) = \sum_{j=0}^{\infty} b_j(\beta) Y_{t-j}$ and $g_j(x;\beta) = E[Z_t(\beta)|X_{t-j} = x]$, $j = 0, \pm 1, \ldots$

$$m_\theta^*(x) = \sum_{j=0}^{\infty} a_j^*(\alpha) g_j(x;\beta)$$

$$\mathcal{H}_\theta(x,y) = -\sum_{j=\pm 1}^{\pm\infty} a_j^+(\alpha) \frac{f_{0,j}(y,x)}{f_0(y) f_0(x)},$$

where $a_j^*(\alpha) = a_j(\alpha)/\sum_{j=0}^{\infty} a_j^2(\alpha)$ and $a_j^+(\alpha) = \sum_{k \neq 0} a_{j+k}(\alpha) a_j(\alpha)/\sum_{l=0}^{\infty} a_l^2(\alpha)$. We now let m vary with θ, that is, (4) is defined for any θ, and let m_θ be the function that minimizes (4); this satisfies $m_\theta = (I - \mathcal{H}_\theta)^{-1} m_\theta^*$ for all θ provided the conditions A1 and A2 hold uniformly over the parameter space. Furthermore, we can define $\theta = \theta_0$ is the minimizer of

$$Q(\theta, m_\theta) = E\left[\left\{Z_t(\beta) - \sum_{j=0}^{\infty} a_j(\alpha) m_\theta(X_{t-j})\right\}^2\right] \tag{9}$$

with respect to $\theta \in \Theta$ and $m_0 = m_{\theta_0}$. We adopt this profiling approach to defining θ_0, m_0 as this is the way our estimation strategy works.

In practice one has to replace m_θ^* and $\mathcal{H}_\theta$ by estimators. Furthermore, one has also to estimate the parameters of the filters A, B.

2.1.2 Further Details

Suppose we have a sample $\{(Y_1, X_1), \ldots, (Y_T, X_T)\}$. The general estimation strategy is

1. For each θ compute estimators of $\widehat{m}_\theta^*, \widehat{\mathcal{H}}_\theta$ of $m_\theta^*, \mathcal{H}_\theta$

2. Solve an empirical version of the equation (7) to obtain an estimator $\widehat{m}_\theta$ of m_θ

3. Choose $\widehat{\theta}$ to minimize the profiled negative log likelihood or least squares criterion with respect to θ. Let $\widehat{m}(x) = \widehat{m}_{\widehat{\theta}}(x)$.

Let $\tau = \tau(T)$ be some truncation parameter and define $Z_t^\tau(\beta) = \sum_{j=0}^{\tau} b_j(\beta) Y_{t-j}$. The choice of truncation depends on the dependence model $A(L), B(L)$. For geometrically declining parameters (as we shall assume) one can work with logarithmic truncation. For long memory sequences it would be necessary to allow for algebraic τ.

For any sequence $\{Z_t^\tau(\beta)\}$ and any lag j define the estimator $\widehat{g}_j(x;\beta) = \widehat{a}_0$, where $(\widehat{a}_0, \widehat{a}_1)$ are the minimizers of the weighted sums of squares criterion

$$\sum_{t=j+1}^{T} \left\{Z_t^\tau(\beta) - a_0 - a_1(X_{t-j} - x)\right\}^2 K_h\left(X_{t-j} - x\right) \tag{10}$$

with respect to (a_0, a_1), where K is a symmetric probability density function, h is a positive bandwidth, and $K_h(.) = K(./h)/h$. Further define

$$\begin{aligned} \widehat{f}_{0,j}(y,x) &= \frac{1}{T-|j|} \sum_{t=|j|+1}^{T} K_h(y - X_t) K_h(x - X_{t-j}), \\ \widehat{f}_0(x) &= \frac{1}{T} \sum_{t=1}^{T} K_h(x - X_t). \end{aligned}$$

$$\widehat{m}_\theta^*(x) = \sum_{j=0}^{\tau} a_j^*(\alpha) \widehat{g}_j(x;\beta)$$

$$\widehat{\mathcal{H}}_\theta(x,y) = -\sum_{j=\pm 1}^{\pm\tau} a_j^+(\alpha) \frac{\widehat{f}_{0,j}(y,x)}{\widehat{f}_0(y)\widehat{f}_0(x)}.$$

Then define $\widehat{m}_\theta$ as any solution to the equation

$$m = \widehat{m}_\theta^* + \widehat{\mathcal{H}}_\theta m, \tag{11}$$

in $L_2(\widehat{f}_0)$. Let $\widehat{\theta} = \arg\min_{\theta \in \Theta} \widehat{Q}_T(\theta)$, where

$$\widehat{Q}_T(\theta) = \frac{1}{T} \sum_{t=2}^{T} \left\{ Z_t^\tau(\beta) - \sum_{j=0}^{\tau} a_j(\alpha) \widehat{m}_\theta(X_{t-j}) \right\}^2.$$

Finally, let $\widehat{m}(x) = \widehat{m}_{\widehat{\theta}}(x)$.

2.2 Asymptotic Properties

We suppose that $\{Y_t, X_t\}$ is a stationary α-mixing process. Let $\mathcal{F}_a^b$ be the σ-algebra of events generated by the random variables $\{Y_t, X_t; \quad a \le j \le b\}$. The stationary processes $\{Y_t, X_t\}$ is called strongly mixing [Rosenblatt (1956)] if

$$\sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} \left|\Pr(A \cap B) - \Pr(A)\Pr(B)\right| \equiv s(k) \to 0 \quad \text{as } k \to \infty. \tag{12}$$

We shall consider two cases. First, the 'weak form case' where we do not maintain that model (1) holds with an i.i.d. error process. Second, we maintain that model (1) holds with a martingale difference error sequence ε_t. To facilitate the asymptotic analysis, we make the following assumptions on the residuals and regressors, the kernel function $k(\cdot)$, and the bandwidth parameter h. Let $\eta_{t,j}(\beta) = Z_t(\beta) - E[Z_t(\beta)|X_{t-j}]$, $\zeta_{t,j}(\theta) = m_\theta(X_t) - E[m_\theta(X_t)|X_{t-j}]$ and

$$\eta^1_{\theta,t} = \sum_{j=1}^{\infty} a^\dagger_j(\alpha)\eta_{t,j}(\beta) \text{ and } \eta^2_{\theta,t} = -\sum_{j=\pm 1}^{\pm\infty} a^*_j(\alpha)\zeta_{t,j}(\theta). \tag{13}$$

B1 *The process* $\{X_t, Y_t\}_{t=-\infty}^{\infty}$ *is stationary and alpha mixing with a mixing coefficient,* $s(k)$ *such that for some* $C \geq 0$ *and some large* s_0, $s(k) \leq Ck^{-s_0}$.

B2 $E\left(|Y_t|^{2\rho}\right) < \infty$ *for some* $\rho > 2$.

B3 *The covariate process* $\{X_t\}_{t=-\infty}^{\infty}$ *has absolutely continuous density* f_0 *supported on* $[-c, c]$ *for some* $c < \infty$. *The function* $m(\cdot)$ *together with the densities* $f_0(\cdot)$, *and* $f_{0,j}(\cdot)$ *are continuous and twice continuously differentiable over* $[-c, c]$, *and are uniformly bounded.* $f_0(\cdot)$ *is bounded away from zero on* $[-c, c]$, *i.e.,* $\inf_{-c\leq w\leq c} f_0(w) > 0$.

B4 *The parameter space* Θ *is a compact subset of* $\mathbb{R}^p$, *and the value* θ_0 *is an interior point of* Θ. *Also, A2 holds, and for any* $\epsilon > 0$

$$\inf_{||\theta-\theta_0||>\epsilon} Q(\theta, m_\theta) > Q(\theta_0, m_{\theta_0}).$$

B5 *The density function* λ of $(\eta^1_{t,j}(\beta), \eta^2_{t,j}(\beta))$ *is Lipschitz continuous on its domain. The joint densities* $\lambda_{0,j}, j = 1, 2, \ldots,$ *of* $(\eta^1_{t,0}(\beta), \eta^2_{t,0}(\beta)), (\eta^1_{t,j}(\beta), \eta^2_{t,j}(\beta))$ *are uniformly bounded.*

B6 *The parameters* $\alpha \in \mathcal{A}$ *and* $\beta \in \mathcal{B}$ *compact subsets of* $\mathbb{R}^{p_a}$ *and* $\mathbb{R}^{p_b}$ *respectively. The coefficients satisfy* $\sup_{\alpha\in A,k=0,1,2} ||\partial^k a_j(\alpha)/\partial\alpha^k|| \leq C\overline{a}^j$ *for some* $\overline{a} < 1$ *and some finite constant* C, *while* $\inf_{\alpha\in\mathcal{A}} \sum_{j=0}^{\infty} a_j^2(\alpha) > 0$. *Likewise,* $\sup_{\beta\in B,k=0,1,2} ||\partial^k b_j(\beta)/\partial\beta^k|| \leq C\overline{b}^j$ *for some* $\overline{b} < 1$ *and some finite constant* C, *while* $\inf_{\beta\in\mathcal{B}} \sum_{j=0}^{\infty} b_j^2(\beta) > 0$.

B7 *The truncation sequence* τ_T *satisfies* $\tau_T = C \log T$ *for some constant* C.

B8 *The bandwidth sequence* $h(T)$ *satisfies* $T^{1/5}h(T) \to \gamma$ *as* $T \to \infty$ *with* γ *bounded away from zero and infinity.*

B9 *The kernel function is a symmetric probability density function with bounded support such that for some constant* C, $|K(u) - K(v)| \leq C|u - v|$. *Define* $\mu_j(K) = \int u^j K(u)du$ *and* $||K||_2^2 = \int u^j K^2(u)du$.

B10 ε_t *satisfies* $E\left[\varepsilon_t|\{X_s\}_{s=-\infty}^{\infty}, \{\varepsilon_{t-j}\}_{j=1}^{\infty}\right] = 0$ *a.s.*

B11 ε_t *is i.i.d. and independent of the process* $\{X_t\}$.

These conditions are similar to Linton and Mammen (2005) but we also need conditions on the $b_j(\beta)$ coefficients and separate conditions on X and Y.

Define the functions $\beta_\theta^j(x)$, $j = 1, 2$, as solutions to the integral equations

$$\beta_\theta^j = \beta_\theta^{*,j} + \mathcal{H}_\theta \beta_\theta^j,$$

in which:

$$\beta_\theta^{*,1}(x) = \frac{\partial^2}{\partial x^2} m_\theta^*(x),$$

$$\beta_\theta^{*,2}(x) = \sum_{j=\pm 1}^{\pm\infty} a_j^*(\theta) \left\{ E(m_\theta(X_{t+j})|X_t = x) \frac{f_0''(x)}{f_0(x)} - \int \left[\nabla_2 f_{0,j}(x,y)\right] \frac{m_\theta(y)}{p_0(x)} dy \right\},$$

where the operator ∇_2 is defined as $\nabla_2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$. Then define

$$\begin{aligned} \omega_\theta(x) &= \frac{||K||_2^2}{f_0(x)} \mathrm{var}[\eta_{\theta,t}^1 + \eta_{\theta,t}^2] \\ b_\theta(x) &= \frac{1}{2}\mu_2(K) \left[\beta_\theta^1(x) + \beta_\theta^2(x)\right], \end{aligned}$$

where $\eta_{\theta,t}^j$, $j = 1, 2$ were defined above in (13). We prove the following theorem in the appendix.

THEOREM 1. *Suppose that B1-B9 hold. Then for each* $\theta \in \Theta$ *and* $x \in (-c, c)$

$$\sqrt{Th}\left[\widehat{m}_\theta(x) - m_\theta(x) - h^2 b_\theta(x)\right] \Longrightarrow N\left(0, \omega_\theta(x)\right), \tag{14}$$

Both the bias and variance in this result are quite complicated even though a local linear smoother has been used in estimating g_j. This is a 'weak form' result, where the model (1) is not assumed.

We next maintain a 'semi-strong form' assumption B10, which allows the filters to be misspecified except that the ensuing error term must still be orthogonal to the covariate process and its own history. We obtain the properties of $\widehat{\theta}$ by an application of the asymptotic theory for semiparametric profiled estimators, see Severini and Wong (1992) and Newey (1994). This requires a uniform expansion for $\widehat{m}_\theta(x)$ and for the derivatives (with respect to θ) of $\widehat{m}_\theta(x)$. Define:

$$\omega(x) = \frac{||K||_2^2 \sum_{j=1}^\infty a_j^2(\alpha_0) E\left[\varepsilon_t^2 | X_{t-j} = x\right]}{f_0(x)\left[\sum_{j=1}^\infty a_j^2(\alpha_0)\right]^2} \tag{15}$$

$$b(x) = \mu_2(K)\left\{\frac{1}{2} m''(x) + (I - \mathcal{H}_\theta)^{-1}\left[\frac{f_0'}{f_0}\frac{\partial}{\partial x}(\mathcal{H}_\theta m)\right](x)\right\}. \tag{16}$$

Let $\varepsilon_t(\theta) = Z_t(\beta) - \sum_{j=0}^\infty a_j(\alpha) m_\theta(X_{t-j})$, and let

$$\mathcal{J} = E\left[\frac{\partial^2 \varepsilon_t}{\partial\theta\partial\theta^\top}(\theta_0)\right] \text{ and } \mathcal{I} = E\left[\frac{\partial \varepsilon_t}{\partial\theta}\frac{\partial \varepsilon_t}{\partial\theta^\top}\varepsilon_t^2(\theta_0)\right].$$

THEOREM 2. *Suppose that Assumptions B1 to B10 hold. Then,*

$$\sqrt{T}(\widehat{\theta}-\theta_0) \Longrightarrow N(0,\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}).$$

Furthermore, for $x \in (-c,c)$

$$\sqrt{Th}\left(\widehat{m}(x)-m(x)-h^2 b(x)\right) \Longrightarrow N\left(0,\omega(x)\right).$$

Note that the autocorrelation of the induced error term ε_t does not affect the limiting variance although its heteroskedasticity does. Under the 'strong form' special case that ε_t is at least conditionally homoskedastic, $\omega(x) = ||K||_2^2\sigma_\varepsilon^2/f_0(x)\sum_{j=0}^{\infty}a_j^2$. Compare this with the usual kernel estimator, which has asymptotic variance $\omega_{Ker}(x) = ||K||_2^2\sigma_\varepsilon^2\sum_{j=0}^{\infty}c_j^2/f_0(x)$, where $C(L) = A(L)^{-1}$. Compare with the estimator of Xiao, Linton, Carroll, and Mammen (2003), which has variance $\omega_{XLCM}(x) = ||K||_2^2\sigma_\varepsilon^2/f_0(x)$. In this case, $\omega(x) \leq \omega_{XLCM}(x) \leq \omega_{Ker}(x)$.

As in Linton and Mammen (2005, p789) it is possible to adjust the operator in order to produce a simpler bias term. The modified estimator has bias

$$b(x) = \left(\lim_{T\to\infty}\sqrt{Th^5}\right)\frac{1}{2}\mu_2(K)m''(x),$$

which is as for a standard local linear estimator in regression. With this implementation then we get a straight mean squared error reduction.

Assumption B10 is needed for the consistency of the parameter estimates $\widehat{\theta}$. In the pure regression model (2, 3) one only needs a weaker assumption $E\left[\varepsilon_t|\{X_s\}_{s=-\infty}^{\infty}\right] = 0$ *a.s.* for consistent estimation of m and θ as is known from the parametric case.

Under the 'strong form' assumption B11 the parametric estimator is semiparametrically efficient.

3 Nonstationary Case

In this section we investigate the case where Y_t can be nonstationary but X_t is stationary mixing as before. The most general case would be where both A, B contained unit roots either simple or complex, so for example $A(L) = (1-L)A'(L)$, where $A'(L)$ obeys the summability conditions in B6. For expositional reason we shall focus on an even more special case where $B(L) = A(L) = 1-L$.

Consider the model

$$(1-\rho L)Y_t = (1-\rho L)m(X_t) + \varepsilon_t, \tag{17}$$

where in fact $\rho_0 = 1$ and ε_t obeys B11. In this case,

$$Y_t = m(X_t) + u_t,$$

where $u_t = u_{t-1} + \varepsilon_t$ is a unit root process, Phillips (1987). We suppose that $u_0 = 0$.

Direct estimation of Y_t on X_t will produce inconsistent estimates of m. On the other hand our estimation of the additive model

$$Y_t - Y_{t-1} = m(X_t) - m(X_{t-1}) + \varepsilon_t$$

with white noise errors will produce consistent estimates of m. In fact, the theory for m_{ρ_0} is exactly as in Theorem 1. The Xiao, Linton, Carroll, and Mammen (2003) procedure is also inconsistent in this unit root case because it relies on the initial standard nonparametric regression estimator that is inconsistent. The task here is to determine that we can estimate the parameter ρ in (17) consistently and thence estimate m consistently.

One issue is that for $\rho \neq 1$, the process $(1-\rho L)Y_t$ is non-stationary and so some of the definitions of the previous section don't make sense. Instead we define $m_{T\rho}$ to be the potentially time varying minimizer of

$$Q_T(m) = \frac{1}{T}\sum_{t=1}^{T} E\left[\{Y_t - \rho Y_{t-1} - m(X_t) + \rho m(X_{t-1})\}^2\right].$$

A necessary condition for m to be the minimizer is that it satisfies the first order condition

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T} E[Y_t - \rho Y_{t-1}|X_t = x] - \rho E[Y_t - \rho Y_{t-1}|X_{t-1} = x] \\ = \ & (1+\rho^2) m_{T\rho}(x) + \rho\left(E[m_{T\rho}(X_t)|X_{t-1} = x] + E[m_{T\rho}(X_{t-1})|X_t = x]\right). \end{aligned} \tag{18}$$

Then note that $Y_t - \rho Y_{t-1} = m(X_t) - \rho m(X_{t-1}) + \varepsilon_t + (1-\rho)u_{t-1}$, and so $E[Y_t - \rho Y_{t-1}|X_t = x]$ and $E[Y_t - \rho Y_{t-1}|X_{t-1} = x]$ are time invariant. Furthermore, we have assumed that X_t is stationary. Therefore, there exists a time invariant solution to equation (18) as in the purely stationary case.[2] Furthermore, the solution is characterized by the integral equation (7) with in this special case:

$$m_\rho^*(x) = \frac{1}{1+\rho^2}\left(E[Y_t - \rho Y_{t-1}|X_t = x] - \rho E[Y_t - \rho Y_{t-1}|X_{t-1} = x]\right)$$

$$\mathcal{H}_\rho(x,y) = -\frac{\rho}{1+\rho^2}\left(\frac{f_{0,1}(y,x)}{f_0(y)f_0(x)} + \frac{f_{0,1}(x,y)}{f_0(y)f_0(x)}\right).$$

What is different here is the error in estimating $E\left[Y_t - \rho Y_{t-1}|X_{t-1} = x\right]$ for example can be large unless ρ is close to one in which case the term $(1-\rho)u_{t-1}$ is small and the process $Y_t - \rho Y_{t-1}$ is almost stationary. The difference in behaviour of the resulting $\widehat{m}_\rho$ for $\rho = 1$ and $\rho \neq 1$ is what drives the faster rate of convergence for $\widehat{\rho}$.

Define

$$\widehat{Q}_T(\rho) = \frac{1}{T}\sum_{t=2}^{T}\{Y_t - \rho Y_{t-1} - \widehat{m}_\rho(X_t) + \rho\widehat{m}_\rho(X_{t-1})\}^2$$

and let $\widehat{\rho} = \arg\min_\rho \widehat{Q}_T(\rho)$. Let B denote the standard Brownian Motion on $[0,1]$.

[2] Note also that $m_\rho = m$ for all ρ.

THEOREM 3. *Suppose that assumption B1 holds for X_t, that B2 holds for ε_t, that B3, B7-B9 and B11 hold. Then*

$$T(\widehat{\rho}-1) \Longrightarrow \frac{\int_0^1 B(s)dB(s)}{\int_0^1 B^2(s)ds}.$$

Furthermore,

$$\sqrt{Th}\left(\widehat{m}(x)-m(x)-h^2b(x)\right) \Longrightarrow N\left(0,\omega(x)\right),$$

where b(x) is defined in (16) and

$$\omega(x) = ||K||_2^2 \frac{E\left[\varepsilon_t^2|X_t=x\right]+E\left[\varepsilon_t^2|X_{t-1}=x\right]}{4f_0(x)}.$$

This can be generalized easily to allow for short run dynamics in addition to the unit root.

4 Extensions

4.1 NonStationary X, Y

Suppose now that

$$X_t = X_{t-1} + \eta_t$$

with η_t also white noise and uncorrelated with ε_t. Thus X_t is a unit root process. This makes a substantial difference to the asymptotics since the operator $\mathcal{H}_\rho(x,y)$ is now random; rates of convergence are slower etc.

5 Numerical Results

We investigate the performance of our procedure on simulated data. We suppose that

$$Y_t = m(X_t) + u_t,\ u_t = \rho_0 u_{t-1} + \varepsilon_t$$

with $m(x) = \beta_0 x^2/2$, where $X_t \sim N(0,1)$, and $\varepsilon_t \sim N(0,\sigma)$. We examine the cases $T \in \{800, 400, 200\}$ and $\rho_0 \in \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$, and use $ns = 1000$ replications. We compute our estimator $\widehat{m}$ using 200 grid points and assuming in the first instance that ρ_0 is known. We also compute the standard local linear estimator $\widetilde{m}$, in both cases the Gaussian kernel was used.

We chose bandwidth to be optimal according to (asymptotic) weighted mean squared error

$$P_\infty(\widehat{m}) = \underset{T\to\infty}{\text{plim}} T^{4/5} \int_{-c}^{c} \left[\widehat{m}(x)-m(x)\right]^2 f_0(x)dx,$$

which gives $h_{opt} = c_K c_M T^{-1/5}$, where $c_K = (2c||K||_2^2/\mu_2^2(K))^{1/5}$ is to do with the kernel and $c_M = (\sigma_\varepsilon^2/(1+\rho_0^2)\beta_0^2(F_0(c)-F_0(-c)))^{1/5}$, where $F_0(x)$ is the c.d.f. of the covariate, is to do with the model.

We have taken $c = 2$, which corresponds to an interval containing almost 95% of the covariate distribution. For the standard local linear estimator the optimal bandwidth is $c_K c_M^* T^{-1/5}$ with $c_M^* = (\sigma_\varepsilon^2/(1-\rho_0^2)\beta_0^2(F_0(c) - F_0(-c))^{1/5}$ provided $\rho_0 \neq 1$ (when $\rho_0 = 1$ we set ρ_0 in the formula arbitrarily to 0.9).

In Figure 1 below we report the relative value of the performance measure

$$P_T(\widehat{m}) = E\int_{-c}^{c} \left[\widehat{m}(x) - m(x)\right]^2 f_0(x)dx$$

to $P_T(\widetilde{m})$, where E is computed by the average over Monte Carlo simulations. Both estimators use their optimal bandwidths, and consequently their theoretical relative efficiency is $((1-\rho_0^2)/(1+\rho_0^2))^{4/5}$. This is plotted below along with the simulation average value for the different sample sizes against ρ values. The results indicate that $\widehat{m}$ is indeed more efficient than $\widetilde{m}$ and that the advantage takes off after $\rho_0 = 0.8$; until this value the advantage is less than 20% in MSE terms. For small values of ρ_0 the finite sample performance ratio is actually better than predicted, although this is partly because $\widetilde{m}$ performs worse than predicted by its asymptotic theory. Note that when $\rho_0 = 1$ the standard local linear estimator is inconsistent.

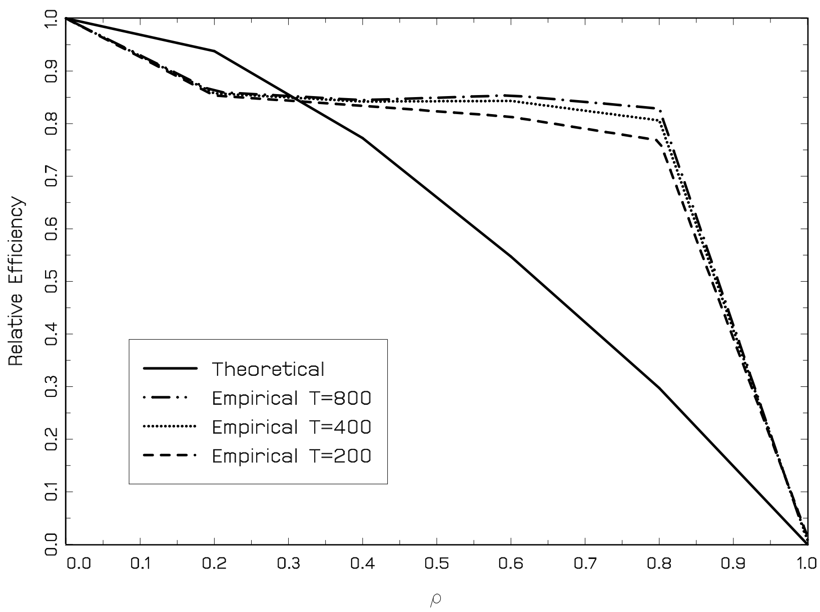

Figure 1. Shows the empirical performance ratio $P_T(\widehat{m})/P_T(\widetilde{m})$ for different sample sizes along with the asymptotic value $P_\infty(\widehat{m})/P_\infty(\widetilde{m})$ predicted from the asymptotic theory. X_t iid $N(0, 1)$.

We also looked at the case where X_t is autocorrelated, specifically, $X_t = 0.95X_{t-1} + u_t$, where u_t is normally distributed such that X_t is marginally $N(0, 1)$. Theoretically, this does not make any difference, and in practice if anything relative performance is improved for this case.

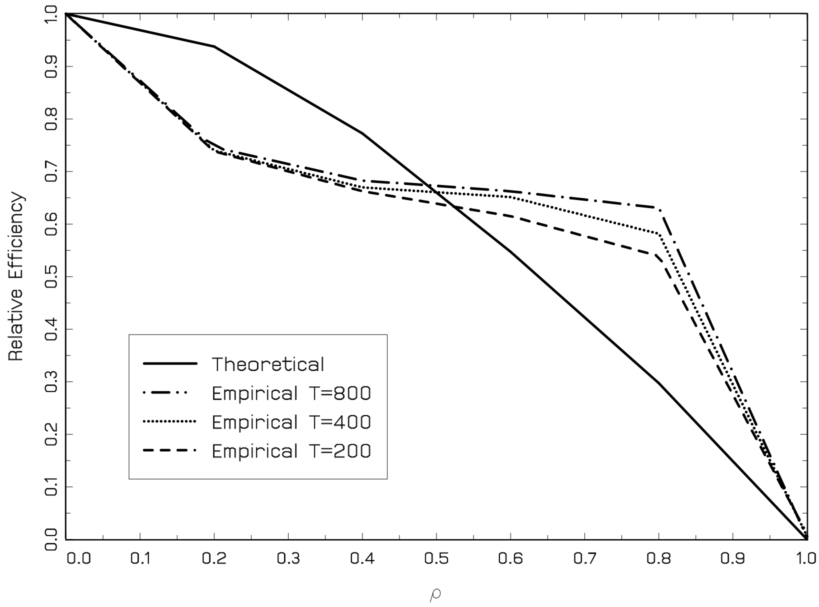

Figure 2. Shows the empirical performance ratio $P_T(\widehat{m})/P_T(\widetilde{m})$ for different sample sizes along with the asymptotic value $P_\infty(\widehat{m})/P_\infty(\widetilde{m})$ predicted from the asymptotic theory. $X_t = 0.95X_{t-1} + u_t$ with $X_t \sim N(0,1)$.

We next examine the performance of the estimates of $\widehat{\rho}$. When $\rho < 1$ these behave pretty much as predicted. When $\rho_0 = 1$, our simulations show that the variance of $\widehat{\rho}$ decreases rapidly with sample size with standard deviation being 0.0161, 0.00896, and 0.00458 for $T = 200, 400$, and 800 respectively. Below we show the densities. As the sample size increases the density approaches the Dicky-Fuller density.

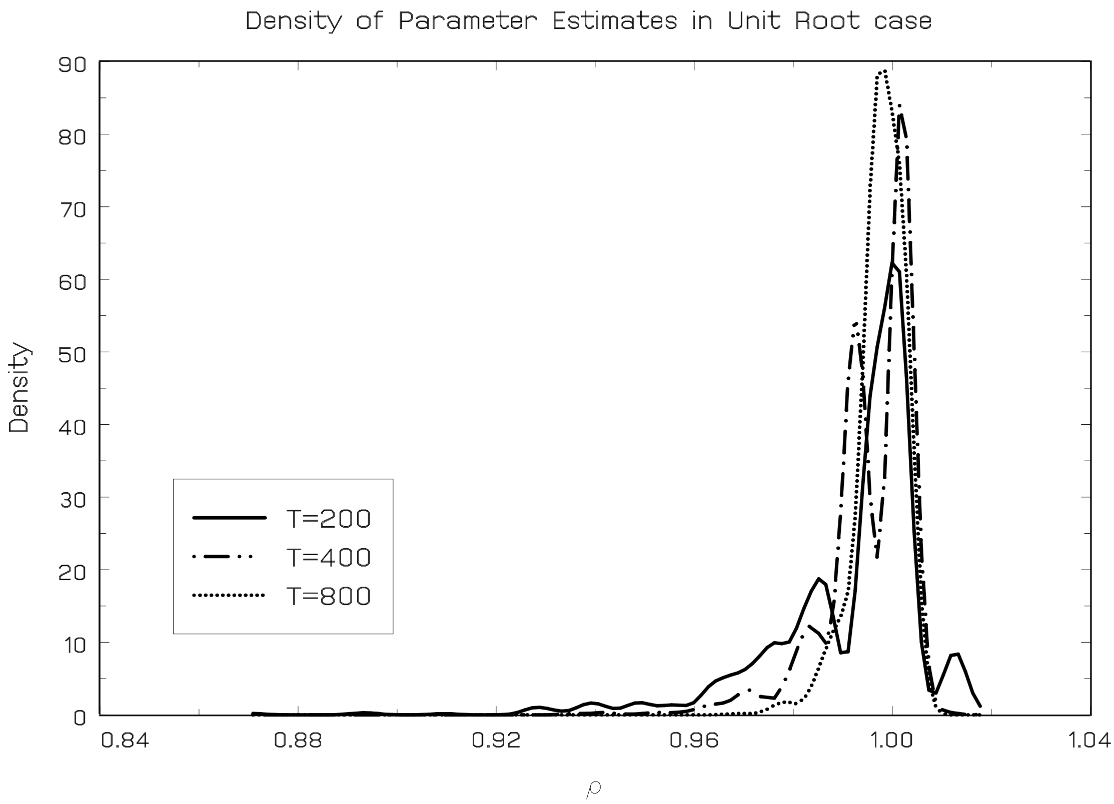

Figure 3. Shows the density of $\widehat{\rho}$ for three diferent sample sizes. $X_t = 0.95X_{t-1} + u_t$ with $X_t \sim N(0,1)$.

A Appendix

A.1 Computational Appendix

We discuss briefly how we solve the equation (11) in practice. Note that one can rewrite (8) as an integral equation on $[0,1]^2$ as $m_\theta^\dagger(s) = m_\theta^{*\dagger}(s) + \int_0^1 \mathcal{H}_\theta^\dagger(s,t)m_\theta(t)dt$, where $\mathcal{H}_\theta^\dagger(s,t) = \mathcal{H}_\theta^\dagger(F_0^{-1}(s), F_0^{-1}(t))$ with $y = F_0^{-1}(s), x = F_0^{-1}(t)$ and $m_\theta^\dagger(t) = m_\theta(F_0^{-1}(t))$ and $m_\theta^{*\dagger}(t) = m_\theta^*(F_0^{-1}(t))$ and F_0 is the c.d.f. of X_t. For simplicity we drop the superfluous $\dagger$ superscript in the sequel. Let $\{t_{j,n}, j = 1,\ldots,n\}$ be some equally spaced grid of points in $[0,1]$, and let $q_{j,n} = \widehat{F}_0^{-1}(t_{j,n})$ be the empirical $t_{j,n}$ quantile of X_t. Now approximate (11) by

$$\widehat{m}_\theta(q_{i,n}) = \widehat{m}_\theta^*(q_{i,n}) + \sum_{j=1}^{n} \widehat{\mathcal{H}}_\theta(q_{i,n}, q_{j,n})\widehat{m}_\theta(q_{j,n}),\ i = 1,\ldots,n. \tag{19}$$

The linear system (19) can be written in matrix notation

$$(I_n - \widehat{\mathbf{H}}_\theta)\widehat{\mathbf{m}}_\theta = \widehat{\mathbf{m}}_\theta^*, \tag{20}$$

where I_n is the $n \times n$ identity, $\widehat{\mathbf{m}}_\theta = (\widehat{m}_\theta(q_{1,n}), \ldots, \widehat{m}_\theta(q_{n,n}))^\top$ and $\widehat{\mathbf{m}}_\theta^* = (\widehat{m}_\theta^*(q_{1,n}), \ldots, \widehat{m}_\theta^*(q_{n,n}))^\top$, while

$$\widehat{\mathbf{H}}_\theta = -\sum_{k=\pm 1}^{\pm\tau} a_k^+(\alpha) \left[\frac{\widehat{f}_{0,k}(q_{i,n}, q_{j,n})}{\widehat{f}_0(q_{i,n}) \widehat{f}_0(q_{j,n})} \right]_{i,j=1}^n$$

is an $n \times n$ matrix. We then find the solution values $\widehat{\mathbf{m}}_\theta = (\widehat{m}_\theta(q_{1,n}), \ldots, \widehat{m}_\theta(q_{n,n}))^\top$ to this system (20) by direct inversion when n is less than say 2000.

A.2 Proof of Theorems

A.2.1 Stationary Case

Proof of Theorem 1. The proof strategy follows Linton and Mammen (2005). First, for general $\theta \neq \theta_0$ we apply Linton and Mammen (2005, Proposition 1). Thus we write

$$\widehat{m}_\theta^*(x) - m_\theta^*(x) = \widehat{m}_\theta^{*,B}(x) + \widehat{m}_\theta^{*,C}(x) + \widehat{m}_\theta^{*,D}(x) \tag{21}$$

$$(\widehat{\mathcal{H}}_\theta - \mathcal{H}_\theta) m_\theta(x) = \widehat{m}_\theta^{*,E}(x) + \widehat{m}_\theta^{*,F}(x) + \widehat{m}_\theta^{*,G}(x), \tag{22}$$

where $\widehat{m}_\theta^{*,B}(x)$ and $\widehat{m}_\theta^{*,E}(x)$ are deterministic and $O(T^{-2/5})$,

$$\begin{aligned}
\widehat{m}_\theta^{*,B}(x) &= \frac{h^2}{2} \mu_2(K) m_\theta^{*\prime\prime}(x) \\
\widehat{m}_\theta^{*,E}(x) &= \frac{h^2}{2} \mu_2(K) \sum_{s=\pm 1}^{\pm\infty} a_j^*(\alpha) \left\{ E(m_\theta(X_{t+j}) | X_t = x) \frac{f_0''(x)}{f_0(x)} - \int \left[\nabla_2 f_{0,j}(x,y) \right] \frac{m_\theta(y)}{f_0(x)} dy \right\}
\end{aligned}$$

where $\nabla_2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$, while

$$\widehat{m}_\theta^{*,C}(x) = \frac{1}{T f_0(x)} \sum_t K_h(X_t - x) \eta_{\theta,t}^1$$

$$\widehat{m}_\theta^{*,F}(x) = \frac{1}{T f_0(x)} \sum_t K_h(X_t - x) \eta_{\theta,t}^2$$

and the remainder terms $\widehat{m}_\theta^{*,D}(x)$ and $\widehat{m}_\theta^{*,G}(x)$ satisfy

$$\sup_{\theta \in \Theta, x \in \mathcal{X}} \left| \widehat{m}_\theta^{*,j}(x) \right| = o_p(T^{-2/5}), \quad j = D, G.$$

From this one obtains an expansion

$$\widehat{m}_\theta(x) - m_\theta(x) = m_\theta^B(x) + m_\theta^E(x) + \widehat{m}_\theta^{*,C}(x) + \widehat{m}_\theta^{*,F}(x) + o_p(T^{-2/5}), \tag{23}$$

where $m_\theta^B = (I - \mathcal{H}_\theta)^{-1} \widehat{m}_\theta^{*,B}$ and $m_\theta^E = (I - \mathcal{H}_\theta)^{-1} \widehat{m}_\theta^{*,E}$, and the error is small uniformly over x.

Step 1. The first step is to establish the expansions (21) and (22). Write

$$Z_t(\beta) - Z_t^\tau(\beta) = \sum_{j=\tau+1}^\infty b_j(\beta) Y_{t-j}.$$

Let $\widetilde{g}_j(x;\beta)$ denote (10) with $Z_t(\beta)$ replacing $Z_t^\tau(\beta)$. Then

$$\max_{1\le j\le\tau}\sup_{x\in\mathcal{X},\beta\in\mathcal{B}}|\widehat{g}_j(x;\beta)-\widetilde{g}_j(x;\beta)|=o_p(T^{-1/2}). \tag{24}$$

This follows because of the assumed decay rates on b_j and the moment condition on Y.

Then for each j,

$$\widetilde{g}_j(x;\beta)-g_j(x;\beta)=\frac{1}{Thf_0(x)}\sum_{t=j+1}^{T}K\left(\frac{x-X_{t-j}}{h}\right)\eta_{t,j}(\beta)+\frac{h^2}{2}\mu_2(K)\mathbf{b}_j(x;\beta)+R_{Tj}(x;\beta),$$

where $\mathbf{b}_j(x;\beta)$ is the bias function and $R_{Tj}(x;\beta)$ is the remainder term. By a change of variables and interchanging the order of summation we obtain

$$\begin{aligned}\sum_{j=0}^{\tau}a_j^*(\alpha)\sum_{t=\tau+1}^{T}K\left(\frac{x-X_{t-j}}{h}\right)\eta_{t-j,j}(\beta) &= \sum_{s=\tau+1}^{T}K\left(\frac{x-X_s}{h}\right)\sum_{j=0}^{\infty}a_j^*(\alpha)\eta_{s,j}(\beta)\\ &\quad-\sum_{s=\tau+1}^{T}K\left(\frac{x-X_s}{h}\right)\sum_{j=\tau+1}^{\infty}a_j^*(\alpha)\eta_{s,j}(\beta)\\ &\quad+\sum_{j=0}^{\tau}a_j^*(\alpha)\sum_{s=\tau+1-j}^{\tau+1}K\left(\frac{x-X_s}{h}\right)\eta_{s,j}(\beta)\\ &\quad+\sum_{j=0}^{\tau}a_j^*(\alpha)\sum_{t=T-j}^{T}K\left(\frac{x-X_s}{h}\right)\eta_{s,j}(\beta),\end{aligned}$$

where the terms apart from the first are of smaller order. Therefore,

$$\begin{aligned}\sum_{j=0}^{\tau}a_j^*(\alpha)[\widehat{g}_j(x;\beta)-g_j(x;\beta)] &= \sum_{j=0}^{\tau}a_j^*(\alpha)\sum_{s=\tau+1}^{T}K\left(\frac{x-X_s}{h}\right)\sum_{j=0}^{\infty}a_j^*(\alpha)\eta_{s,j}(\beta)\\ &\quad+\frac{h^2}{2}\mu_2(K)\sum_{j=0}^{\tau}a_j^*(\alpha)\mathbf{b}_j(x;\beta)+o_p(T^{-2/5})\end{aligned}$$

Note that uniformly over $j\le\tau$ and over x,

$$\frac{1}{Th}\sum_{t=j+1}^{T}K\left(\frac{x-X_{t-j}}{h}\right)-\frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{x-X_t}{h}\right)=O_p(\tau/T)=o_p(T^{-1/2}),$$

so that one can shift the indexes with impunity.

We next establish the expansion (22). We have

$$\begin{aligned}&\int\widehat{\mathcal{H}}_\theta(x,y)m_\theta(y)\widehat{f}_0(y)dy-\int\mathcal{H}_\theta(x,y)m_\theta(y)f_0(y)dy\\ &= -\sum_{j=\pm1}^{\pm\tau}a_j^{+}(\alpha)\int\left[\frac{\widehat{f}_{0,j}(x,y)}{\widehat{f}_0(x)}-\frac{f_{0,j}(x,y)}{f_0(x)}\right]m_\theta(y)dy.\end{aligned}$$

Denote by

$$\int \frac{f_{0,j}(x,y)}{f_0(x)} m_\theta(y)dy = E\left[m(X_{t-j})|X_t = x\right] \equiv r_j(x).$$

Then write

$$\frac{\int \widehat{f}_{0,j}(x,y)m_\theta(y)dy}{\widehat{f}_0(x)} = \frac{\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)m^*_{t-j}}{\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)}, \tag{25}$$

where

$$m^*_{t-j} = \frac{1}{h}\int K\left(\frac{y-X_{t-j}}{h}\right)m_\theta(y)dy = \int K(u)m_\theta(X_{t-j}+uh)du \simeq m_\theta(X_{t-j}) + \frac{h^2}{2}\mu_2(K)m''_\theta(X_{t-j}).$$

This is just like a local constant smoother of m^*_{t-j} on X_t and can be analyzed in the same way.

Therefore using $\widehat{a}/\widehat{b} - c = (\widehat{a} - \widehat{b}c)/\widehat{b}$ we have

$$\begin{aligned}
&\frac{\int \widehat{f}_{0,j}(x,y)m_\theta(y)dy}{\widehat{f}_0(x)} - \int \frac{f_{0,j}(x,y)}{f_0(x)}m_\theta(y)dy \\
= \;& \frac{\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)\left[m^*_{t-j} - r_j(x)\right]}{\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)} \\
= \;& \frac{\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)\left[m_\theta(X_{t-j}) - r_j(x)\right]}{\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)} + \frac{\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)\left[m^*_{t-j} - m_\theta(X_{t-j})\right]}{\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)} \qquad (26)\\
\simeq \;& \frac{\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)\left[m_\theta(X_{t-j}) - r_j(X_t)\right]}{\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)} + \frac{\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)\left[r_j(X_t) - r_j(x)\right]}{\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)} \\
&+ \frac{h^2}{2}\mu_2(K)E[m''_\theta(X_{t-j})|X_t = x] \\
\simeq \;& \frac{1}{Th}\frac{1}{f_0(x)}\sum_t K\left(\frac{x-X_t}{h}\right)\zeta_{t,j} + \frac{h^2}{2}\mu_2(K)\left[r''_j(x) + \frac{2r'_j(x)f'_0(x)}{f_0(x)} + E[m''_\theta(X_{t-j})|X_t = x]\right] \qquad (27)
\end{aligned}$$

by standard arguments where $\zeta_{t,j} = m_\theta(X_{t-j}) - r_j(X_t) = m_\theta(X_{t-j}) - E[m_\theta(X_{t-j})|X_t]$ is mean zero sequence given X_t.

We have

$$\begin{aligned}
E[m''_\theta(X_{t-j})|X_t = x] &= \int \frac{f_{0,j}(x,y)}{f_0(x)}m''_\theta(y)dy = \int \frac{\partial^2 f_{0,j}(x,y)/\partial y^2}{f_0(x)}m_\theta(y)dy \\
r'_j(x) &= \frac{1}{f_0(x)}\int \frac{\partial f_{0,j}(x,y)}{\partial x}m_\theta(y)dy - \frac{f'_0(x)}{f_0^2(x)}\int f_{0,j}(x,y)m_\theta(y)dy \\
r''_j(x) &= \frac{1}{f_0(x)}\int \frac{\partial^2 f_{0,j}(x,y)}{\partial x^2}m_\theta(y)dy + \left[\frac{2(f'_0(x))^2}{f_0^3(x)} - \frac{f''_0(x)}{f_0^2(x)}\right]\int f_{0,j}(x,y)m_\theta(y)dy \\
&\quad -2\frac{f'_0(x)}{f_0^2(x)}\int \frac{\partial f_{0,j}(x,y)}{\partial x}m_\theta(y)dy.
\end{aligned}$$

The bias terms in (27) are

$$\frac{h^2}{2}\mu_2(K)\left[r_j''(x)+\frac{2r_j'(x)f_0'(x)}{f_0(x)}+\frac{1}{f_0(x)}\int\frac{\partial^2 f_{0,j}(x,y)}{\partial y^2}m_\theta(y)dy\right]. \tag{28}$$

However, there is a cancellation

$$\begin{aligned}
&2\frac{f_0'(x)}{f_0^2(x)}\int\frac{\partial f_{0,j}(x,y)}{\partial x}m_\theta(y)dy-2\frac{(f_0'(x))^2}{f_0^3(x)}\int f_{0,j}(x,y)m_\theta(y)dy+\frac{1}{f_0(x)}\int\frac{\partial^2 f_{0,j}(x,y)}{\partial x^2}m_\theta(y)dy\\
&+\left[\frac{2(f_0'(x))^2}{f_0^3(x)}-\frac{f_0''(x)}{f_0^2(x)}\right]\int f_{0,j}(x,y)m_\theta(y)dy\\
&-2\frac{f_0'(x)}{f_0^2(x)}\int\frac{\partial f_{0,j}(x,y)}{\partial x}m_\theta(y)dy\\
=\ &\frac{1}{f_0(x)}\int\frac{\partial^2 f_{0,j}(x,y)}{\partial x^2}m_\theta(y)dy-\frac{f_0''(x)}{f_0^2(x)}\int f_{0,j}(x,y)m_\theta(y)dy,
\end{aligned}$$

so the bias (28) is

$$\begin{aligned}
&\frac{h^2}{2}\mu_2(K)\left[\frac{1}{f_0(x)}\int\frac{\partial^2 f_{0,j}(x,y)}{\partial x^2}m_\theta(y)dy-\frac{f_0''(x)}{f_0^2(x)}\int f_{0,j}(x,y)m_\theta(y)dy+\frac{1}{f_0(x)}\int\frac{\partial^2 f_{0,j}(x,y)}{\partial y^2}m_\theta(y)dy\right]\\
=\ &\frac{h^2}{2}\mu_2(K)\left[\frac{1}{f_0(x)}\int\frac{\partial^2 f_{0,j}(x,y)}{\partial x^2}m_\theta(y)dy-\frac{f_0''(x)}{f_0(x)}r_j(x)+\frac{1}{f_0(x)}\int\frac{\partial^2 f_{0,j}(x,y)}{\partial y^2}m_\theta(y)dy\right].
\end{aligned}$$

In conclusion, we have

$$\begin{aligned}
&\int\widehat{\mathcal{H}}_\theta(x,y)m_\theta(y)\widehat{f}_0(y)dy-\int\mathcal{H}_\theta(x,y)m_\theta(y)f_0(y)dy\\
=\ &-\frac{1}{f_0(x)}\frac{1}{Th}\sum_t K\left(\frac{x-X_t}{h}\right)\sum_{j=\pm1}^{\pm\tau}a_j^{\dagger}(\alpha)\zeta_{t,j}\\
&+\frac{h^2}{2}\mu_2(K)\sum_{j=\pm1}^{\pm\tau}a_j^{\dagger}(\alpha)\left[\frac{f_0''(x)}{f_0(x)}r_j(x)-\frac{1}{f_0(x)}\int\left(\frac{\partial^2 f_{0,j}(x,y)}{\partial x^2}+\frac{\partial^2 f_{0,j}(x,y)}{\partial y^2}\right)m_\theta(y)dy\right]\\
&+o_p(T^{-2/5}).
\end{aligned}$$

STEP 2. The $\sqrt{T}$-consistency of $\widehat{\theta}$ follows along the lines of Linton and Mammen (2005) using the expansions obtained above uniform over θ.

STEP 3. This implies that one can treat θ as known and one obtains a simpler expansion for $\widehat{m}_{\theta_0}(x)-m(x)$. In particular:

$$\eta_{t,j}=Z_{t+j}-E[Z_{t+j}|X_t]=\varepsilon_{t+j}+\sum_{k\neq j}a_k\left[m(X_{t+j-k})-E[m(X_{t+j-k})|X_t]\right]$$

$$\sum_{j=1}^{\infty}a_j^{\dagger}\eta_{t,j}=\sum_{j=1}^{\infty}a_j^{\dagger}\varepsilon_{t+j}+\sum_{j=1}^{\infty}a_j^{\dagger}\sum_{k\neq j}a_k\left[m(X_{t+j-k})-E[m(X_{t+j-k})|X_t]\right]$$

$$\sum_{j=1}^{\infty} a_j^{\dagger} \sum_{k \neq j} a_k \left[m(X_{t+j-k}) - E[m(X_{t+j-k})|X_t] \right] = \sum_{j=\pm 1}^{\pm\infty} a_j^{+} \left[m(X_{t+j}) - E[m(X_{t+j})|X_t] \right] = \sum_{j=\pm 1}^{\pm\infty} a_j^{+} \zeta_{t,j},$$

where $a_j^{\dagger} = a_j / \sum_{j=0}^{\infty} a_j^2$ and $a_j^{+} = \sum_{k \neq 0} a_{j+k} a_j / \sum_{j=0}^{\infty} a_j^2$. It follows that

$$\sum_{j=1}^{\infty} a_j^{\dagger} \eta_{t,j} - \sum_{j=\pm 1}^{\pm\infty} a_j^{+} \zeta_{t,j} = \sum_{j=1}^{\infty} a_j^{\dagger} \varepsilon_{t-j}. \tag{29}$$

Therefore, the stochastic part of (23), $\widehat{m}_{\theta}^{*,C}(x) + \widehat{m}_{\theta}^{*,F}(x)$, simplifies. Likewise, there is a simplification for the bias term $m_{\theta}^{B}(x) + m_{\theta}^{E}(x)$.

■

A.2.2 Nonstationary Case

PROOF OF THEOREM 2. Let

$$\varepsilon_t(\rho) = Y_t - \rho Y_{t-1} - m_\rho(X_t) + \rho m_\rho(X_{t-1}) = Y_t - \rho Y_{t-1} - m(X_t) + \rho m(X_{t-1}) = \varepsilon_t + (1-\rho)u_{t-1}$$

$$\widehat{\varepsilon}_t(\rho) = Y_t - \rho Y_{t-1} - \widehat{m}_\rho(X_t) + \rho \widehat{m}_\rho(X_{t-1})$$

$$Q_T(\rho) = \frac{1}{T}\sum_{t=2}^{T} \varepsilon_t^2(\rho).$$

We first establish the properties of an estimator that minimizes $Q_T(\rho)$, denoted $\overline{\rho}$. In our case,

$$\begin{aligned} Q_T(\rho) &= \frac{1}{T}\sum_{t=2}^{T}\varepsilon_t^2 + T(1-\rho)^2\frac{1}{T^2}\sum_{t=2}^{T}u_{t-1}^2 + 2(1-\rho)\frac{1}{T}\sum_{t=2}^{T}\varepsilon_t u_{t-1} \\ &\simeq \sigma_\varepsilon^2 + T(1-\rho)^2\sigma_\varepsilon^2\int B^2(s)ds + 2(1-\rho)\sigma_\varepsilon^2\int B(s)dB(s), \end{aligned}$$

where B is the standard Brownian motion, from which we obtain consistency of $\overline{\rho}$ at rate T and furthermore

$$T(\overline{\rho}-1) \Longrightarrow \frac{\int B(s)dB(s)}{\int B^2(s)ds}.$$

We next consider the difference between $\widehat{Q}_T(\rho)$ and $Q_T(\rho)$. We have

$$\widehat{Q}_T(\rho) = Q_T(\rho) + \frac{1}{T}\sum_{t=2}^{T}\left\{\widehat{\varepsilon}_t(\rho) - \varepsilon_t(\rho)\right\}^2 + 2\frac{1}{T}\sum_{t=2}^{T}\left\{\widehat{\varepsilon}_t(\rho) - \varepsilon_t(\rho)\right\}\varepsilon_t(\rho), \tag{30}$$

$$\widehat{\varepsilon}_t(\rho) - \varepsilon_t(\rho) = -(\widehat{m}_\rho(X_t) - m_\rho(X_t)) + \rho(\widehat{m}_\rho(X_{t-1}) - m_\rho(X_{t-1})).$$

STEP 1. When $\rho = 1$ the properties of $\widehat{Q}_T(\rho)$ can be derived as in the stationary case. In particular,

$$\widehat{Q}_T(1) \to^p q \tag{31}$$

for some $q > 0$; hence, $\widehat{Q}_T(1)/T \to^p 0$.

STEP 2. Derive the properties of $\widehat{m}_\rho - m_\rho$ for $\rho \neq 1$. As in the stationary case we can approximate $\widehat{m}_\rho - m_\rho$ in terms of $\widehat{m}_\rho^* - m_\rho^*$ and $(\widehat{\mathcal{H}}_\rho - \mathcal{H}_\rho)m_\rho$. The expansion for $(\widehat{\mathcal{H}}_\rho - \mathcal{H}_\rho)m_\rho$ is as above. The main difference concerns the fact that the expansion for $\widehat{m}_\rho^* - m_\rho^*$ contains a term that is large when $\rho \neq 1$ and indeed $\widehat{m}_\rho^*$ does not consistently estimate m_ρ^* unless $\rho = 1$. Therefore, $\widehat{m}_\rho - m_\rho$ is dominated by the large term in $\widehat{m}_\rho^* - m_\rho^*$. The intercept function m_ρ^* is

$$\begin{aligned} m_\rho^*(x) &= \frac{1}{1+\rho^2}\left(E[Y_t - \rho Y_{t-1}|X_t = x] - \rho E[Y_t - \rho Y_{t-1}|X_{t-1} = x]\right) \\ &= \frac{1}{1+\rho^2}\left[g_{0\rho}(x) - \rho g_{1\rho}(x)\right], \end{aligned}$$

a linear combination of $g_{0\rho}(x) = E[Y_t - \rho Y_{t-1}|X_t = x]$ and $g_{1\rho}(x) = E[Y_t - \rho Y_{t-1}|X_{t-1} = x]$. Therefore, we must establish the properties of $\widehat{g}_{j\rho}(x) - g_{j\rho}(x)$, $j = 0, 1$, where $\widehat{g}_{j\rho}(x)$ are the estimates of $g_{j\rho}(x)$.

STEP 3. Derive the properties of $\widehat{g}_{j\rho}(x) - g_{j\rho}(x)$, $j = 0, 1$ and $\rho \neq 1$. We have

$$Y_t - \rho Y_{t-1} - E\left[Y_t - \rho Y_{t-1}|X_t = x\right] = m(X_t) - m(x) - \rho\left(m(X_{t-1}) - E[m(X_{t-1})|X_t = x]\right) + \varepsilon_t + (1-\rho)u_{t-1}.$$

$$Y_t - \rho Y_{t-1} - E\left[Y_t - \rho Y_{t-1}|X_{t-1} = x\right] = m(X_t) - E[m(X_t)|X_{t-1} = x] - \rho\left(m(X_{t-1}) - m(x)\right) + \varepsilon_t + (1-\rho)u_{t-1}.$$

The terms $m(X_t) - m(x)$ and $m(X_{t-1}) - m(x)$ on the rhs contribute to biases; the stationary error terms $-\rho\left(m(X_{t-1}) - E[m(X_{t-1})|X_t = x]\right) + \varepsilon_t$ and $m(X_t) - E[m(X_t)|X_{t-1} = x] + \varepsilon_t$ may contribute to the variance but are standard, it is the term $(1-\rho)u_{t-1}$ containing the unit root that is different. We have

$$\begin{aligned} \widehat{g}_{j\rho}(x) - g_{j\rho}(x) &= \frac{1}{Thf_0(x)} \sum_{t=j+1}^{T} K\left(\frac{x - X_{t-j}}{h}\right)\varepsilon_t + (1-\rho)\frac{1}{Thf_0(x)} \sum_{t=j+1}^{T} K\left(\frac{x - X_{t-j}}{h}\right)u_{t-1} \\ &\quad + \frac{h^2}{2}\mu_2(K)\mathbf{b}_j(x;\rho) + R_T(x;\rho) \equiv \delta_{T1} + \delta_{T2} + \delta_{T3} + R_T(x;\rho), \end{aligned}$$

where the remainder term $R_T(x;\rho)$ is of smaller order. This approximation is valid because the X process is stationary so everything except δ_{T2} is standard. We consider the term δ_{T2} and write $\delta_{T2} = \sqrt{T}(1-\rho)\xi_T(x) + \sqrt{T}(1-\rho)\eta_T(x)$ with

$$\begin{aligned} \xi_T(x) &= \frac{1}{T}\sum_{t=1}^{T} E\left[\frac{1}{hf_0(x)}K\left(\frac{x - X_{t-j}}{h}\right)\right]\frac{u_{t-1}}{\sqrt{T}} \\ \eta_T(x) &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{hf_0(x)}K\left(\frac{x - X_{t-j}}{h}\right) - E\left[\frac{1}{hf_0(x)}K\left(\frac{x - X_{t-j}}{h}\right)\right]\right)\frac{u_{t-1}}{\sqrt{T}}. \end{aligned}$$

Clearly,

$$\xi_T(x) = \frac{1}{T}\sum_{t=1}^{T}\frac{u_{t-1}}{\sqrt{T}} + o_p(1) = O_p(1)$$

for all x.

We argue that $\eta_T(x) = o_p(1)$. Note that $E[\eta_T(x)] = 0$ by independence of X, u processes. Define

$$\epsilon_{Tt} = \frac{1}{hf_0(x)}K\left(\frac{x - X_{t-j}}{h}\right) - E\left[\frac{1}{hf_0(x)}K\left(\frac{x - X_{t-j}}{h}\right)\right]. \qquad (32)$$

This has (approximately as $T \to \infty$) covariance function

$$\begin{aligned} \mathrm{cov}(\epsilon_{Tt}, \epsilon_{Tt-r}) &\simeq E\left[\frac{1}{h^2 f_0^2(x)}K\left(\frac{x - X_t}{h}\right)K\left(\frac{x - X_{t-r}}{h}\right)\right] - E^2\left[\frac{1}{hf_0(x)}K\left(\frac{x - X_t}{h}\right)\right] \\ &\simeq \frac{f_{0,t-r}(x,x)}{f_0^2(x)} - 1 \equiv \gamma_\epsilon(t-r). \end{aligned}$$

Furthermore,

$$\begin{aligned} \text{var}\left[\eta_T(x)\right] &= \frac{1}{T^3}\sum_{t=j+1}^{T} E\left[\epsilon_{Tt}^2\right] E[u_{t-1}^2] + \frac{1}{T^3}\sum\sum_{t\neq s} E\left[\epsilon_{Tt}\epsilon_{Ts}\right] E[u_t u_s] \\ &\simeq \frac{\sigma_\varepsilon^2}{T^3}\sum\sum_{t\neq s}\min\{s,t\}\gamma_\epsilon(t-s). \end{aligned}$$

We have

$$\frac{1}{T^3}\sum\sum_{t\neq s}\min\{s,t\}\gamma_\epsilon(t-s) \simeq \frac{2}{T^3}\sum_{s=1}^{T-1} s\sum_{t=s+1}^{T}\gamma_\epsilon(t-s) \simeq \frac{2}{T^2}\sum_{s=1}^{T-1} s(T-s)\sum_{k=1}^{\infty}\gamma_\epsilon(k) = \frac{2}{3T}\sum_{k=1}^{\infty}\gamma_\epsilon(k)$$

so that $\text{var}\left[\eta_T(x)\right] = O(T^{-1})$. Therefore

$$\widehat{g}_{j\rho}(x) - g_{j\rho}(x) \simeq \sqrt{T}(1-\rho)\frac{1}{T}\sum_{t=1}^{T}\frac{u_{t-1}}{\sqrt{T}},$$

which is the same regardless of location x and j. By the usual arguments (Phillips (1987))

$$\frac{1}{T}\sum_{t=1}^{T}\frac{u_{t-1}}{\sqrt{T}} \Longrightarrow \sigma_\varepsilon\int_0^1 B(s)ds. \tag{33}$$

Therefore, $(\widehat{g}_{j\rho}(x) - g_{j\rho}(x))/\sqrt{T} \Longrightarrow (1-\rho)\sigma_\varepsilon\int_0^1 B(s)ds$ for all x.

STEP 4. Obtain an approximation to $\widehat{\varepsilon}_t(\rho) - \varepsilon_t(\rho)$. For $\rho \neq 1$ we have

$$\begin{aligned} &\widehat{\varepsilon}_t(\rho) - \varepsilon_t(\rho) \\ =\ & -(\widehat{m}_\rho(X_t) - m(X_t)) + \rho(\widehat{m}_\rho(X_{t-1}) - m(X_{t-1})) \\ \simeq\ & -(\widehat{m}_\rho^* - m_\rho^*)(X_t) + \rho(\widehat{m}_\rho^* - m_\rho^*)(X_{t-1}) \\ \simeq\ & \frac{-1}{1+\rho^2}\left[[\widehat{g}_{0\rho} - g_{0\rho}](X_t) + \rho^2[\widehat{g}_{1\rho} - g_{1\rho}](X_{t-1}) - \rho[\widehat{g}_{0\rho} - g_{0\rho}](X_{t-1}) - \rho[\widehat{g}_{1\rho} - g_{1\rho}](X_t)\right] \\ \simeq\ & \frac{-(1-\rho)^3}{1+\rho^2}\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\frac{u_{t-1}}{\sqrt{T}}, \end{aligned}$$

because the other terms are of smaller order in probability.

STEP 5. Obtain properties of the terms in $\widehat{Q}_T(\rho)$.

1. We have
$$\frac{1}{T}\sum_{t=2}^{T}\{\widehat{\varepsilon}_t(\rho) - \varepsilon_t(\rho)\}^2 \simeq \frac{(1-\rho)^6 T}{(1+\rho^2)^2}\sigma_\varepsilon^2\left(\int_0^1 B(s)ds\right)^2.$$

2. We have
$$(1-\rho)\frac{1}{T}\sum_{t=2}^{T}\{\widehat{\varepsilon}_t(\rho) - \varepsilon_t(\rho)\}u_{t-1} \simeq \frac{-(1-\rho)^4 T}{1+\rho^2}\sigma_\varepsilon^2\left(\int_0^1 B(s)ds\right)^2.$$

3. We have

$$\frac{1}{T}\sum_{t=2}^{T}\{\widehat{\varepsilon}_t(\rho)-\varepsilon_t(\rho)\}\,\varepsilon_t \simeq \frac{-(1-\rho)^3}{1+\rho^2}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\frac{u_{t-1}}{\sqrt{T}}\varepsilon_t = O_p(1).$$

STEP 6. Obtain an expansion for $\widehat{Q}_T(\rho)$. We have

$$\begin{aligned}\widehat{Q}_T(\rho) \;\; &\simeq \;\; \sigma_\varepsilon^2 + T(1-\rho)^2\sigma_\varepsilon^2\int B^2(s)ds + 2(1-\rho)\sigma_\varepsilon^2\int B(s)dB(s) \\ &+\frac{(1-\rho)^6 T}{(1+\rho^2)^2}\sigma_\varepsilon^2\left(\int_0^1 B(s)ds\right)^2 - \frac{2(1-\rho)^4 T}{1+\rho^2}\sigma_\varepsilon^2\left(\int_0^1 B(s)ds\right)^2 - \frac{-2(1-\rho)^3}{1+\rho^2}O_p(1).\end{aligned}$$

Therefore

$$\begin{aligned}\frac{1}{T}\widehat{Q}_T(\rho) \;\; &\simeq \;\; (1-\rho)^2\sigma_\varepsilon^2\int B^2(s)ds + \left[\frac{(1-\rho)^6}{(1+\rho^2)^2} - \frac{2(1-\rho)^4}{1+\rho^2}\right]\sigma_\varepsilon^2\left(\int_0^1 B(s)ds\right)^2 \\ &= \;\; (1-\rho)^2\sigma_\varepsilon^2\int B^2(s)ds - \frac{(1-\rho)^4(\rho+1)^2}{(1+\rho^2)^2}\sigma_\varepsilon^2\left(\int_0^1 B(s)ds\right)^2.\end{aligned}$$

By Cauchy-Schwarz

$$\int_0^1 B^2(s)ds \geq \left(\int_0^1 B(s)ds\right)^2.$$

Therefore, with probability one:

$$\frac{1}{T}\widehat{Q}_T(\rho) \geq 4\,(1-\rho)^2\,\frac{\rho^2}{(1+\rho^2)^2}\sigma_\varepsilon^2\left(\int_0^1 B(s)ds\right)^2 \geq 0 \tag{34}$$

for all $\rho \neq 1$. Hence, with probability tending to one

$$\lim_{T\to\infty}\inf_{|\rho-1|>\delta}\frac{1}{T}\widehat{Q}_T(\rho) > 0. \tag{35}$$

STEP 7. Combine (31) and (35) yields $\widehat{\rho} \xrightarrow{P} 1$, see for example Pakes and Pollard (1989).

STEP 8. Then reparameterizing $\rho \mapsto r = 1-\rho/T$ we get

$$\widehat{Q}_T(r) \simeq \sigma_\varepsilon^2 + \frac{r^2}{T}\sigma_\varepsilon^2\int B^2(s)ds + 2\frac{r}{T}\sigma_\varepsilon^2\int B(s)dB(s) + o(T^{-1})$$

so that the asymptotic distribution is just the Dickey-Fuller

$$T(\widehat{\rho}-1) \Longrightarrow \frac{\int B(s)dB(s)}{\int B^2(s)ds}.$$

■

References

Andersen, T.W. (1971). *The Statistical Analysis of Time Series.* New York: John Wiley and Sons.

Carrasco, M., J.P. Florens, and E. Renault (2002), "Linear Inverse problems in Structural Econometrics," Forthcoming in *Handbook of Econometrics*, volume 6, eds. J.J. Heckman and E. Leamer.

Conley, T.G., L.P. Hansen, E.G.J. Luttmer, and J.A. Scheinkman (1997), "Short-Term Interest Rates as Subordinated Diffusions," *The Review of Financial Studies,* 10, 525-577.

Fan, J. (1992). Design-Adaptive Nonparametric Regression. Journal of the American Statistical Association, 87 , 998-1004.

Fan, J. and I. Gijbels (1996). Local Polynomial Modelling and Its Applications, Champman & Hall.

Hannan, E.J., and M. Deistler (1988). *The Statistical Theory of Linear Systems.* John Wiley; New York.

Hart, J.D. (1991). Kernel Regression Estimation with Time Series Errrors. *Journal of the Royal Statistical Society.* 53, 173-187.

Harvey, A.C. (1981). *The Econometric Analysis of Time Series.* Philip Alan, Oxford.

Hastie, T. and R. Tibshirani (1991). *Generalized Additive Models.* Chapman and Hall, London.

Hidalgo, F.J. (1997): "Non-parametric estimation with strongly dependent multivariate time series," *Journal of Time Series Analysis* 18, 95-122.

Lin, X. and Carroll, R. J. (2000), "Nonparametric Function Estimation for Clustered Data When the Predictor is Measured Without/With Error," Journal of the American Statistical Association, 95, 520-534.

Linton, O.B. (1997). Efficient estimation of additive nonparametric regression models. *Biometrika*, **84**, 469-474.

Linton, O.B., and E. Mammen (2005): "Estimating semiparametric ARCH(∞) models by kernel smoothing methods**,"** *Econometrica* 73, 771-836.

Marinucci, D., and P.M. Robinson (2001): "Semiparametric fractional cointegration analysis," STICERD WM/01/420.

MASRY, E. (1996a). Multivariate regression estimation: Local polynomial fitting for time series. *Stochastic Processes and their Applications.* **65**, 81-101.

MASRY, E. (1996b). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *J. Time Ser. Anal.* **17**, 571-599.

MULLER, H., AND U. STADTMULLER (1988), Detecting dependencies in smooth regression models, *Biometrika*, 75, 4, pp 639-650.

OPSOMER, J., Y. WANG, AND Y. YANG (2001). Nonparametric Regression with Correlated Errors. Manuscript.

PHILLIPS, P.C.B. (1987): "Time Series Regression with a Unit Root," *Econometrica* 55, 277-301.

PHILLIPS, P.C.B., AND J.Y. PARK (1998): "Nonstationary density estimaiton and kernel autoregression," CFDP no.

ROBINSON, P.M. (1983). Nonparametric Estimators for Time Series. Journal of Time Series Analysis 4, 185-207.

ROSENBLATT, M. (1956). A central limit theorem and strong mixing conditions, *Proc. Nat. Acad. Sci.* **4**, 43-47.

RUCKSTUHL, A., WELSH, A. H. AND CARROLL, R. J. (2000), "Nonparametric Function Estimation of the Relationship Between Two Repeatedly Measured Variables," Statistica Sinica, 10, 51–71.

VILAR-FERNANDEZ, J.M. AND M. FRANCISCO-FERNANDEZ (2001). Local polynomial regression smoothers with AR-error structure. Forthcoming in TEST.

XIAO, Z., O. LINTON, R. J. CARROLL, AND E. MAMMEN (2003). More Efficient Local Polynomial Estimation in Nonparametric Regression with Autocorrelated Errors *Journal of the American Statistical Association* 98, 980-992.